

CLAIMS

WHAT IS CLAIMED IS:

Sub B

1. A method for generating a decision tree for a plurality of data characterized by a plurality of features, comprising:

- 3 selecting a feature from among the features;
- 4 performing a cluster analysis along the selected feature to group the data into one or
5 more clusters; and
- 6 building the decision tree based on the one or more clusters.

- 1 2. The method according to claim 1, wherein the step of selecting the feature
2 includes the steps of:

- 3 3. performing a plurality of cluster analyses along each of the features to calculate a
4 maximal cluster validity measure, said maximal cluster validity measure
5 corresponding to one of the features; and
- 6 4. selecting the one of the features that corresponds to the maximal cluster validity
7 measure.

- 1 5. The method according to claim 2, wherein the step of performing a plurality of
2 cluster analyses along each of the features to calculate a maximal cluster validity measure
3 includes the performing the steps of:

- 4 6. for each of the features, performing a plurality of cluster analyses along said each of
5 the features for a plurality of cluster numbers to calculate respective partition
6 coefficients; and
- 7 7. determining the maximal cluster validity measure from among the partition
8 coefficients.

- 1 4. The method according to claim 1, wherein the step of performing the cluster
- 2 analysis includes the step of performing a fuzzy cluster analysis.

- 1 5. The method according to claim 4, wherein the step of performing the fuzzy cluster
- 2 analysis includes the step of performing a fuzzy c-means analysis.

- 1 6. The method according to claim 1, wherein the step of performing the cluster
- 2 analysis includes the step of performing a hard cluster analysis.

- 1 7. The method according to claim 1, wherein the step of performing the cluster
- 2 analysis along the selected feature to group the data into one or more clusters includes the
- 3 steps of:
 - 4 calculating a domain ratio of a difference in domains limits of the data over a
 - 5 difference in domain limits of a superset of the data;
 - 6 determining whether the domain ratio has a predetermined relationship with a
 - 7 predetermined threshold; and
 - 8 if the domain ratio has the predetermined relationship with the predetermined
 - 9 threshold, then grouping the data into a single cluster.

- 1 8. The method according to claim 7, wherein the step of determining whether the
- 2 domain ratio has the predetermined relationship with the predetermined threshold
- 3 includes the step of determining whether the domain ratio is less than the predetermined
- 4 threshold.

1 9. The method according to claim 1, wherein building the decision tree based on the
2 one or more clusters includes the steps of:

3 projecting the data in each of the clusters, wherein the projected data are
4 characterized by the plurality of the features but for the selected feature; and
5 recursively performing the steps of selecting a feature and performing the cluster
6 analysis on the projected data in each of the clusters.

1 10. A method for generating a decision tree for a plurality of data characterized by a
2 plurality of features, comprising:

3 performing a plurality of cluster analyses along each of the features to calculate a
4 maximal cluster validity measure, said maximal cluster validity measure
5 corresponding to one of the features;
6 selecting the one of the features corresponding to the maximal cluster validity
7 measure;
8 subdividing the data into one or more groups based on the selected feature; and
9 building the decision tree based on the one or more groups.

1 11. The method according to claim 10, wherein the step of performing the cluster
2 analyses along each of the features to calculate a maximal cluster validity measure
3 includes the performing the steps of:

4 for each of the features, performing a plurality of cluster analyses along said each of
5 the features for a plurality of cluster numbers to calculate respective partition
6 coefficients; and
7 determining the maximal cluster validity measure from among the partition
8 coefficients.

1 12. The method according to claim 10, wherein the step of performing the cluster
2 analyses includes the step of performing a plurality of fuzzy cluster analyses.

1 13. The method according to claim 10, wherein the step of performing the fuzzy
2 cluster analyses includes the step of performing a plurality of fuzzy c-means analyses.

1 14. The method according to claim 10, wherein the step of performing the cluster
2 analyses includes the step of performing a plurality of hard cluster analyses.

1 15. The method according to claim 10, wherein the step of performing the cluster
2 analyses includes the steps of:

3 calculating a domain ratio of a difference in domains limits of the data over a
4 difference in domain limits of a superset of the data;

5 determining whether the domain ratio has a predetermined relationship with a
6 predetermined threshold; and

7 if the domain ratio has the predetermined relationship with the predetermined
8 threshold, then grouping the data into a single cluster.

1 16. The method according to claim 10, wherein building the decision tree based on
2 the one or more groups includes the steps of:

3 projecting the data in each of the groups, wherein the projected data are characterized
4 by the plurality of the features but for the selected feature; and

5 recursively performing the steps of selecting a feature, comprising selecting a new
6 one of the features corresponding to a new maximal partition coefficient and
7 subdividing the data into one or more new groups based on the selected new
8 feature.

1 17. A method for generating a decision tree for a plurality of data characterized by a
2 plurality of features, comprising:
3 performing a plurality of fuzzy cluster analyses along each of the features to calculate
4 a maximal partition coefficient and a corresponding set of one or more fuzzy
5 clusters, said maximal partition coefficient corresponding to one of the features;
6 selecting the one of the features corresponding to the maximal partition coefficient;
7 and
8 building the decision tree based on the corresponding set of one or more fuzzy
9 clusters.

1 18. A computer-readable medium bearing instructions for generating a decision tree
2 for a plurality of data characterized by a plurality of features, said instructions being
3 arranged to cause one or more processors upon execution thereby to perform the steps of:
4 selecting a feature from among the features;
5 performing a cluster analysis along the selected feature to group the data into one or
6 more clusters; and
7 building the decision tree based on the one or more clusters.

1 19. The computer-readable medium according to claim 18, wherein the step of
2 selecting the feature includes the steps of:
3 performing a plurality of cluster analyses along each of the features to calculate a
4 maximal cluster validity measure, said maximal cluster validity measure
5 corresponding to one of the features; and
6 selecting the one of the features that corresponds to the maximal cluster validity
7 measure.

1 20. The computer-readable medium according to claim 19, wherein the step of
2 performing a plurality of cluster analyses along each of the features to calculate a
3 maximal cluster validity measure includes the performing the steps of:
4 for each of the features, performing a plurality of cluster analyses along each of
5 the features for a plurality of cluster numbers to calculate respective partition
6 coefficients; and
7 determining the maximal cluster validity measure from among the partition
8 coefficients.

1 21. The computer-readable medium according to claim 18, wherein the step of
2 performing the cluster analysis includes the step of performing a fuzzy cluster analysis.

1 22. The computer-readable medium according to claim 21, wherein the step of
2 performing the fuzzy cluster analysis includes the step of performing a fuzzy c-means
3 analysis.

1 23. The computer-readable medium according to claim 18, wherein the step of
2 performing the cluster analysis includes the step of performing a hard cluster analysis.

1 24. The computer-readable medium according to claim 18, wherein the step of
2 performing the cluster analysis along the selected feature to group the data into one or
3 more clusters includes the steps of:
4 calculating a domain ratio of a difference in domains limits of the data over a
5 difference in domain limits of a superset of the data;
6 determining whether the domain ratio has a predetermined relationship with a
7 predetermined threshold; and

8 if the domain ratio has the predetermined relationship with the predetermined
9 threshold, then grouping the data into a single cluster.

1 25. The computer-readable medium according to claim 24, wherein the step of
2 determining whether the domain ratio has the predetermined relationship with the
3 predetermined threshold includes the step of determining whether the domain ratio is less
4 than the predetermined threshold.

1 26. The computer-readable medium according to claim 18, wherein building the
2 decision tree based on the one or more clusters includes the steps of:
3 projecting the data in each of the clusters, wherein the projected data are
4 characterized by the plurality of the features but for the selected feature; and
5 recursively performing the steps of selecting a feature and performing the cluster
6 analysis on the projected data in each of the clusters.

1 27. A computer-readable medium bearing instructions for generating a decision tree
2 for a plurality of data characterized by a plurality of features, said instructions being
3 arranged to cause one or more processors upon execution thereby to perform the steps of:
4 performing a plurality of cluster analyses along each of the features to calculate a
5 maximal cluster validity measure, said maximal cluster validity measure
6 corresponding to one of the features;
7 selecting the one of the features corresponding to the maximal cluster validity
8 measure;
9 subdividing the data into one or more groups based on the selected feature; and
10 building the decision tree based on the one or more groups.

1 28. The computer-readable medium according to claim 27, wherein the step of
2 performing the cluster analyses along each of the features to calculate a maximal cluster
3 validity measure includes the performing the steps of:

4 for each of the features, performing a plurality of cluster analyses along said each of
5 the features for a plurality of cluster numbers to calculate respective partition
6 coefficients; and
7 determining the maximal cluster validity measure from among the partition
8 coefficients.

1 29. The computer-readable medium according to claim 27, wherein the step of
2 performing the cluster analyses includes the step of performing a plurality of fuzzy
3 cluster analyses.

1 30. The computer-readable medium according to claim 27, wherein the step of
2 performing the fuzzy cluster analyses includes the step of performing a plurality of fuzzy
3 c-means analyses.

1 31. The computer-readable medium according to claim 27, wherein the step of
2 performing the cluster analyses includes the step of performing a plurality of hard cluster
3 analyses.

1 32. The computer-readable medium according to claim 27, wherein the step of
2 performing the cluster analyses includes the steps of:
3 calculating a domain ratio of a difference in domains limits of the data over a
4 difference in domain limits of a superset of the data;
5 determining whether the domain ratio has a predetermined relationship with a
6 predetermined threshold; and

7 if the domain ratio has the predetermined relationship with the predetermined
8 threshold, then grouping the data into a single cluster.

1 33. The computer-readable medium according to claim 27, wherein building the
2 decision tree based on the one or more groups includes the steps of:
3 projecting the data in each of the groups, wherein the projected data are characterized
4 by the plurality of the features but for the selected feature; and
5 recursively performing the steps of selecting a feature, comprising selecting a new
6 one of the features corresponding to a new maximal partition coefficient and
7 subdividing the data into one or more new groups based on the selected new
8 feature.

1 34. A computer-readable medium bearing instructions for generating a decision tree
2 for a plurality of data characterized by a plurality of features, said instructions being
3 arranged to cause one or more processors upon execution thereby to perform the steps of:
4 performing a plurality of fuzzy cluster analyses along each of the features to calculate
5 a maximal partition coefficient and a corresponding set of one or more fuzzy
6 clusters, said maximal partition coefficient corresponding to one of the features;
7 selecting the one of the features corresponding to the maximal partition coefficient;
8 and
9 building the decision tree based on the corresponding set of one or more fuzzy
10 clusters.